

# Comparison of machine learning models for breast cancer diagnosis

**Rania R. Kadhim, Mohammed Y. Kamil**

College of Sciences, Mustansiriyah University, Baghdad, Iraq

## Article Info

### Article history:

Received Feb 5, 2022

Revised Jul 1, 2022

Accepted Jul 30, 2022

### Keywords:

Breast cancer

Classification

Computer-aided diagnosis

Machine learning

Wisconsin

## ABSTRACT

Breast cancer is the most common cause of death among women worldwide. Breast cancer can be detected early, and the death rate can be reduced. Machine learning (ML) techniques are a hot topic for study and have proved influential in cancer prediction and early diagnosis. This study's objective is to predict and diagnose breast cancer using ML models and evaluate the most effective based on six criteria: specificity, sensitivity, precision, accuracy, F1-score and receiver operating characteristic curve. All work is done in the anaconda environment, which uses Python's NumPy and SciPy numerical and scientific libraries, and pandas and matplotlib. This study used the Wisconsin diagnostic breast cancer dataset to test ten ML algorithms: decision tree, linear discriminant analysis, forests of randomized trees, gradient boosting, passive aggressive, logistic regression, naïve Bayes, nearest centroid, support vector machine, and perceptron. After collecting the findings, we performed a performance evaluation and compared these various classification techniques. Gradient boosting model outperformed all other algorithms, scoring 96.77% on the F1-score.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



## Corresponding Author:

Mohammed Y. Kamil

College of Sciences, Mustansiriyah University

Baghdad, Iraq

Email: m80y98@uomustansiriyah.edu.iq

## 1. INTRODUCTION

Breast cancer is a significant cause of mortality among women, and it is characterized by an unchecked growth in the number of irregular cells in the breast. Unfortunately, stopping the growth of breast tumors has proven impossible until now [1]. Early detection of breast cancers is critical for accurate diagnosis, and this disease is responsible for about 15% of all cancer-related deaths [2]. Therefore, breast tissue lumps should be detected as early as possible to maximize the chances of survival. In the early stages, tumors can be treated and provide valuable information regarding the emergence of cancer, which can assist boost survival rates [3]. Breast biopsy, mammography, ultrasound, thermography, and fine-needle aspiration cytology are just a few of the procedures that can be used to diagnose this condition. To now, mammography has been widely used for detecting breast cancer; in some cases, this is insufficient, and doctors will require a biopsy before making a final decision [4]. Furthermore, the accuracy of the detection rate is low 60% and 70%. Additional tests, some of them expensive and time-consuming, will be required of the patient [5].

Breast cancer detection uses mammograms [6], ultrasounds, and biopsies, as well as machine learning (ML) system testing, to find out if someone has the disease [7]. The term biopsies are used to take a sample of breast cells for testing. Breast cancer can be diagnosed with a biopsy, a reliable procedure. A biopsy test directs using a specialized needle instrument to obtain a sample of questionable tissue during a biopsy. Biopsy may be asked to have a small metal marker placed in the breast to aid future imaging tests.

Experts in a laboratory examine the cells to see if they are cancerous. It's also important to know what type of cells are involved in breast cancer, how aggressive it is (grade), and if the cells have hormones or other receptors that could alter your treatment options [8]. After a breast biopsy, the medical decision may take a few days. As soon as a biopsy is completed, the breast tissue is sent to the specialized laboratory, then analyzed by an expert in the field of blood and body tissue research (pathologist). The pathologist's report includes information about the tissue samples' size, consistency, and location, whether cancerous or non-cancerous cells. At times, these experts have different viewpoints, which leads to more surgery necessary to harvest additional tissue for further location analysis [9]. Researchers have used ML approaches to overcome the drawbacks of typical diagnostic methods, provide a second opinion to physicians, and lower the likelihood of human error, which could result in patient death [10]. In decision support systems, ML techniques are beneficial in increasing the survival rate because they aid in the automation of the decision-making process. Improve precision and speed of response increases and reduce the amount of work that doctors have to do, especially when there is a medical personnel shortage [11]. Decision-support and health-monitoring systems that use ML approaches can help improve other medical services, such as preventive care and mental health treatments for specific individuals [12].

This research examined strategies for detecting breast cancer using ML algorithms and Wisconsin diagnostic breast cancer (WDBC) dataset: Azar and El-Metwally (2013) [13] used three different types of decision tree algorithms. Decision trees are classified into three categories: boosted decision tree (BDT), single decision tree (SDT), and decision tree forest (DTF). BDT outperformed SDT across the board. The BDT had a sensitivity, accuracy specificity, precision, and area under curve (AUC) of 98.42%, 98.83%, 99.58%, 99.77%, and 0.99, respectively. Senapati and Dash (2013) [14] implemented an artificial neural networks (ANN) algorithm for categorizing breast cancer tumors and suggested the use of the firefly algorithm (FA) to enhance the performance of a local linear wavelet neural network. Specifically, FA was used to improve the parameters of a local linear wavelet neural network. The approach achieved a classification accuracy of 98.14%. Azar and El-Said (2014) [15] described a type of support vector machine (SVM) algorithms; they could rapidly, easily, and efficiently diagnose breast cancer. According to the training data, Lagrangian SVM has the lowest accuracy 95.6107%, whereas standard SVM beat all other techniques in terms of accuracy 97.71%, sensitivity 98.24%, specificity 95.08%, precision 97.39%, and AUC 0.99. Aličković and Subasi (2017) [16] presented data mining techniques with and without genetic algorithm-based feature selection to appropriately categorize medical data. The method simplifies computing and speeds up data mining. In the second stage, to choose between two groups. Individual and multiple algorithm methods were used to develop an accurate system for detecting breast cancer. The accuracy of the random forest is 99.48%, and AUC is 0.999. Wang *et al.* (2018) [17] presented twelve different SVM that are hybridized based on weighted area under the curve ensemble (WAUCE) method. The model performance has been measured; the experimental findings indicated the model was achieved the highest accuracy 97.10%, sensitivity 97.11%, specificity 97.23%, and AUC 0.97. Kumar *et al.* (2020) [18] focused on implementing several classification approaches for data mining and classifying tumors as malignant or benign breast cancer. These twelve algorithms had a combined score of 94%. With an accuracy of 73.21%, only Naive Bayes has underperformed the others. Algorithms based on random trees and random forests have performed remarkably well, with accuracy close to 99%. And the sensitivity, which was found to be 99.10%. Lahoura *et al.* (2021) [19] employed ANN as a algorithm in a system for cloud-based breast cancer detection. Comparisons were made between the best extreme learning machine performance results obtained from stand-alone systems and cloud solutions. As a result of the tests' accuracy is 98.68%, precision is 90.54%, sensitivity is 91.30%, and the F1-score is 81.29%.

The main contribution of this paper is to utilize the capability of ML models and implement them in breast cancer diagnosis. The best performance results of ML are compared by finding optimum values for its weights. Additionally, almost all models available are compared for finding the optimal, simple, and efficient model can breast cancer diagnosis. The rest of this paper is categorized: in section 2, a methodology is presented; in section 3, the experimental environment's setup is described; in section 4, the different results obtained are discussed in this work; and in section 5, we state the importance of the results, conclusions and future work.

## 2. METHOD

The goal of this work was to determine the most effective and predictive algorithm for breast cancer diagnosis. We applied ML algorithms decision tree, linear discriminant analysis, forests of randomized trees, gradient boosting (GB), passive aggressive (PA), logistic regression (LR), Naïve Bayes (NB), nearest centroid (NC), and SVM, perceptron to accomplish on the WDBC dataset. Also, we evaluated the results to determine which model provides the highest accuracy. The methodology begins with data acquisition and

continues with pre-processing, including attribute selection. The pre-processed data is employed to improve ML algorithms that predict breast cancer. To assess the algorithms' performance, the model takes new data that has been labeled-typically accomplished by dividing the labeled data collected into two parts. First part, split 80% of used data to train ML model, which is referred to as the training data or training set. While 20% of the data is utilized to evaluate the model's performance; this is called test data or test set. After evaluating the models, we compare the findings to determine which algorithm gives the highest accuracy and which is the most predictive for breast cancer detection.

Predictive analysis of ML techniques is accomplished by using these algorithms: A decision tree (DT) is a supervised learning non-parametric approach for classification. The DT consists of many nodes that form a rooted tree, which is a tree with no incoming edges and a "root" node. One incoming edge connects each of the other nodes. An internal or test node is a node having outgoing edges, and all additional nodes are referred to as leaves (decision or terminal nodes). DT is easy to understand and interpret. It's possible to see trees in your mind's eye. It is able to work with both numerical and category information [20]. A linear discriminant analysis (LDA) is a well-known algorithm featuring linear decision surfaces. This algorithm is appealing because they offer closed-form solutions that are simple to calculate, are intrinsically multiclass, have been demonstrated to function in reality, and have no hyperparameters to tune [21]. A Forests of randomized trees (FRT) are from ensemble learning methods for resolving complex problems by merging several algorithms based on DT predictions. The FRT algorithm decides the outcome by averaging the output of different trees. As the number of trees grows, the result's precision improves. FRT produces precise predictions that are simple to learn. It can successfully handle vast amounts of data. The FRT algorithm is more accurate in predicting outcomes than the DT method [22]. A GB refers to an extension of boosting to arbitrary differentiable loss functions. GB is a ML approach to improve a model's predictive value by continuously improving its predictive value. Many weak learners are combined into one powerful learner in GB. Weak learners used are the decision trees for each user. Although learning a boosting algorithm can be time-consuming, it is also quite accurate. When it comes to learning algorithms, slower-learning models outperform faster learners. Each iteration of the DT changes the values of the coefficients, weights, or biases assigned to each of the input variables used to estimate the target value to lower the loss function (the measure of the difference between the predicted and actual target values) [23]. A PA is a collection of ML techniques. However, it can be quite beneficial and effective in some situations. PA algorithms are commonly employed for large-scale learning [24]. It is highly beneficial in cases where there is a large quantity of data, and training the whole dataset is computationally impossible due to the sheer volume of the data [25]. A LR is a robust ML technique for classifying incoming data and generating probabilities from discrete and continuous datasets. LR can quickly identify the most efficient factors for classification and categorize observations based on various data sources [26]. LR is used instead of fitting a regression line and fitting an "S" shaped logistic function in LR, which predicts two maximum values (1 or 0). The probability shows the curve of the logistic function [27]. A NB is a simple algorithm to build and is highly successful when dealing with large data. NB is known to outperform even the most powerful classification because it is simple and requires fewer data [28]. Determining the class of the test data set is simple and easy when the independence threshold is reached [29]. The NC algorithm is an algorithm that assigns labels to a class of training samples based on their centroid (mean) distance from the observed object or data. The distances between the empirical data and various class centroids are sorted, and the closest distance is chosen [30]. Perceptron is a kind of neural network, and it is used for binary classification (two-class) as a linear classification approach. It learns a decision boundary in the feature space that uses a line to separate two classes (hyperplane). It is excellent for algorithms that can be readily divided using a linear model and is well suited for large-scale learning [31]. SVM is a ML algorithm that can generalize across two different classes if the method is provided with a set of labeled data in the training set. For the support vector classifier (SVC), the most important task is finding a hyperplane that can distinguish between similar and different data classes. This model is effective because it can work well in complex three-dimensional environments. The method is still effective even when the number of dimensions exceeds the number of samples [32].

### 3. EXPERIMENT ENVIRONMENT

The WDBC dataset was collected by Walberg *et al.* from the University of Wisconsin Hospitals in Madison [33]. WDBC may be found in the University of California Irvine repository for ML [18]. A data of a fine needle aspirate (FNA) of a breast lump is used to compute features; cell nuclei are described in terms of the image's features. WDBC was the patient ID, 30 tumor characteristics, and one class indicator included in the WDBC dataset, containing 569 cases (62.74% benign, 37.26% malignant) [19]. There were 30 characteristics in total, including area, texture, radius, perimeter, compactness, smoothness, concavity, concave points, and symmetry. The ID number feature was removed because it has no impact on classification. The scikit-learn library and the Python language were used for all of the experiments on the

ML algorithms (DT, LDA, FRT, GB, LR, PA, Perceptron, NB, NC, and SVC) discussed in this work. It was implemented with Python's NumPy, SciPy, pandas, and matplotlib libraries.

#### 4. RESULTS AND DISCUSSION

WDBC was classified into benign and malignant tumors using ten classification methods: DT, LDA, FRT, GB, LR, PA, Perceptron, NB, NC, and SVC. For each model, we evaluated the performance of models from five aspects: specificity, sensitivity, precision, accuracy, F1-score. Figure 1 provides the results obtained after the models implemented specificity. This result is significant at the maximum of 100% for LDA, LR, PA, NB, and SVC models. Figure 2 presents the results obtained of sensitivity. It can be seen that the Perceptron model had the highest value of 97.87%.

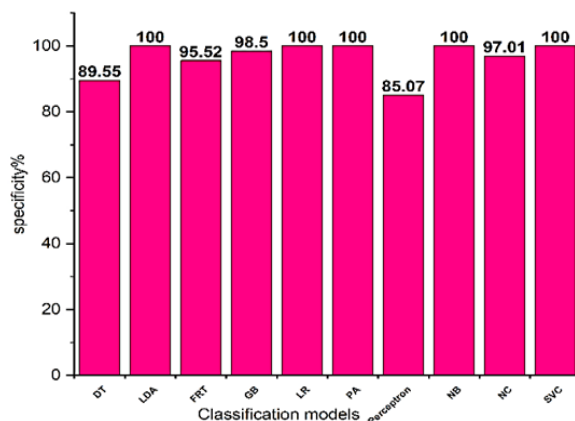


Figure 1. The specificity of classification models

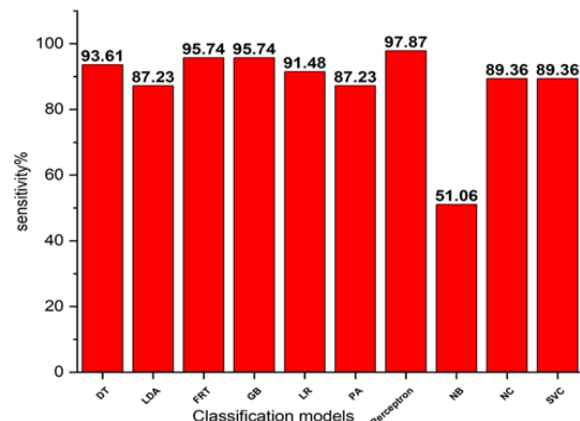


Figure 2. The sensitivity of classification models

The results of the accuracy analysis can be compared in Figure 3 for all used models. There was a significant positive accuracy among almost all models; the results indicate that the GB model had the highest accuracy of 97.36%. Figure 4 shows the results obtained for the precision of models. LDA, LR, PA, NB, and SVC models had values of 100%, whilst the precision for the perceptron model was the lowest at 82.14%. The average scores of specificity and sensitivity can be compared using the F1-score as shown in Figure 5. It is apparent that the GB model had the highest value of 96.77%. The most notable result to observe from the results is the efficiency of ML in the prediction and diagnosis of breast cancer. The mean accuracy was 94.43% for all implemented models except the NB model. We observed GB model achieved the highest results with this data, was stable, and its results were high in all respects (F1-score mainly).

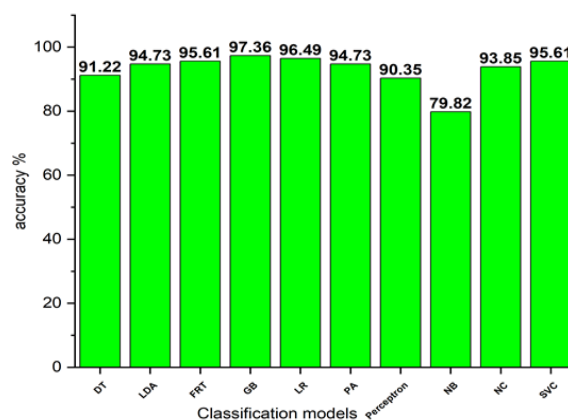


Figure 3. The accuracy of classification models

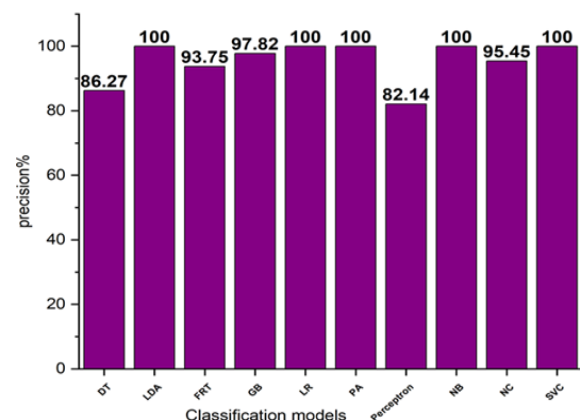


Figure 4. The precision of classification models

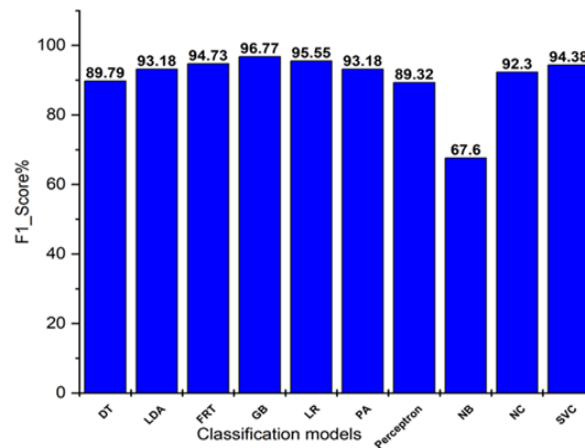


Figure 5. F1-score for classification models

Furthermore, radiologists are interested in the outcome of the trade-off between sensitivity and specificity. Hence, sensitivity refers to a diagnostic's ability to classify patients with cancer as abnormal. In contrast, specificity refers to a diagnostic's ability to classify patients who do not have cancer as normal. As for ML programmers, they are focused on the receiver operating characteristic (ROC) curve results. The ROC curve is created by computing and graphing the true positive rate vs. the false positive rate. Since it shows the model's stability and reliability, it reached a maximum value of 0.99 in GB, as illustrated in the Figure 6.

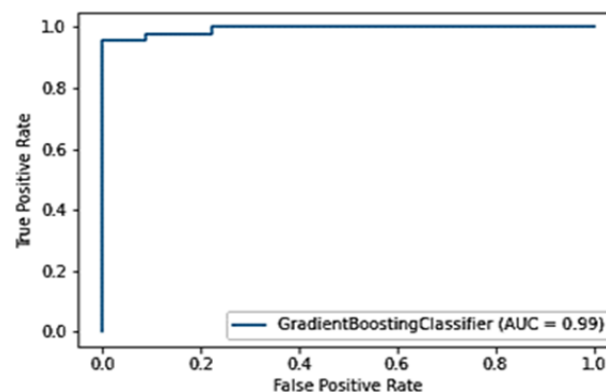


Figure 6. ROC curve for gradient boosting

This result may be explained by the fact that GB is an ensemble method that learns from current predictor errors to improve future predictions. The technique combines the weak classification model into one powerful learner, improving model predictability. By putting weak learners in sequence, weak learners learn from the next learner, improving prediction models. Ensemble methods are suitable for minimizing model variation and improving prediction accuracy. The variance is decreased when many models are joined to generate a single prediction. New predictors are fitted to overcome earlier predictor faults. GB increases the model's accuracy by adding predictors sequentially to the ensemble. GB uses the gradient to discover and correct errors in learners' predictions.

The best results achieved in this work are compared with other results discussed in the literature are highlighted in Table 1. These results agree with the findings of other studies in which ML effectively predicts the presence of breast cancer. To conclude, the ML models accomplished a higher classification accuracy rate, decreased the false positive, and obtained a higher performance rate. As a result of this research, it has been shown that the ML is beneficial in assisting a radiologist in making an accurate breast cancer diagnosis and that it may form the basis for future work that may eventually be used to assist radiologists.

Table 1. Comparison of performance with previous researchers' results

Reference	Year	Machine learning technique	Accuracy %	Sensitivity %	Specificity %	Precision %	AUC
Azar and El-Metwally [13]	2013	BDT	98.83	98.42	99.58	99.77	0.99
Senapati and Dash [14]	2013	wavelet-ANN	98.14	-	-	-	-
Azar and El-Said [15]	2014	LPSVM	97.14	98.24	95.08	97.39	0.99
Aličković and Subasi [16]	2017	Random forest	99.48	-	-	-	0.99
Wang <i>et al.</i> [17]	2018	SVM	97.10	97.11	97.23	-	0.97
		WAUCE					
Kumar <i>et al.</i> [18]	2020	FRT	99.14	99.10	-	99.10	-
		DT	99.14	99.10	-	99.10	-
Lahoura <i>et al.</i> [19]	2021	ANN	98.68	91.30	90.54	81.29	-
Our work		GB	97.36	97.87	<b>100</b>	<b>100</b>	0.99

## 5. CONCLUSION

The early identification of breast cancer is one of the most important scientific topics being pursued. The purpose of the research was to compare the performance of classification algorithms. The efficiency of ML in breast cancer prediction and diagnosis has been recognized to be the most important outcome to emerge from these findings. It is clear from this research that the GB algorithm outperforms with WDBC in terms of specificity and precision. Because this model is part of the ensemble approaches group, we anticipate that the high results will be ascribed to the fact that employing ensembles rather than a single predictive model increases predictive modeling performance in this case. These challenges will need to be resolved in more future studies. For this reason, future research will pay close attention to assessing ML models in additional medical diagnostic issues and optimizing their performance by using high-speed computing methods in the process.

## ACKNOWLEDGEMENTS

The authors would like to thank Mustansiriyah University for their valuable support and providing the essential facilities for this research.




## REFERENCES

- [1] A. H. Farhan and M. Y. Kamil, "Texture analysis of breast cancer via LBP, HOG, and GLCM techniques," *IOP Conference Series: Materials Science and Engineering*, vol. 928, no. 7, 2020, doi: 10.1088/1757-899X/928/7/072098.
- [2] M. Y. Kamil, "Computer-aided diagnosis system for breast cancer based on the Gabor filter technique," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 5, pp. 5235–5242, 2020, doi: 10.11591/IJECE.V10I5.PP5235-5242.
- [3] P. N. Pandey, N. Saini, N. Sapre, D. A. Kulkarni, and D. A. K. Tiwari, "Prioritising breast cancer theranostics: a current medical longing in oncology," *Cancer Treatment and Research Communications*, vol. 29, 2021, doi: 10.1016/j.ctarc.2021.100465.
- [4] B. A. Zeidan, P. A. Townsend, S. D. Garbis, E. Copson, and R. I. Cutress, "Clinical proteomics and breast cancer," *Surgeon*, vol. 13, no. 5, pp. 271–278, 2015, doi: 10.1016/j.surge.2014.12.003.
- [5] M. H. Alshayegi, H. Ellethy, S. Abed, and R. Gupta, "Computer-aided detection of breast cancer on the Wisconsin dataset: an artificial neural networks approach," *Biomedical Signal Processing and Control*, vol. 71, 2022, doi: 10.1016/j.bspc.2021.103141.
- [6] A. H. Farhan and M. Y. Kamil, "Texture analysis of mammogram using local binary pattern method," *Journal of Physics: Conference Series*, vol. 1530, no. 1, 2020, doi: 10.1088/1742-6596/1530/1/012091.
- [7] I. Steinberg, D. M. Huland, O. Vermesh, H. E. Frostig, W. S. Tummers, and S. S. Gambhir, "Photoacoustic clinical imaging," *Photoacoustics*, vol. 14, pp. 77–98, 2019, doi: 10.1016/j.pacs.2019.05.001.
- [8] S. Alimirzaie, M. Bagherzadeh, and M. R. Akbari, "Liquid biopsy in breast cancer: A comprehensive review," *Clinical Genetics*, vol. 95, no. 6, pp. 643–660, 2019, doi: 10.1111/cge.13514.
- [9] L. Wang, "Early diagnosis of breast cancer," *Sensors (Switzerland)*, vol. 17, no. 7, 2017, doi: 10.3390/s17071572.
- [10] E. P. V. Le, Y. Wang, Y. Huang, S. Hickman, and F. J. Gilbert, "Artificial intelligence in breast imaging," *Clinical Radiology*, vol. 74, no. 5, pp. 357–366, 2019, doi: 10.1016/j.crad.2019.02.006.
- [11] E. A. Radhi and M. Y. Kamil, "Breast tumor detection via active contour technique," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 4, pp. 561–570, 2021, doi: 10.22266/ijies2021.0831.49.
- [12] N. I. R. Yassin, S. Omran, E. M. F. El Houbby, and H. Allam, "Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review," *Computer Methods and Programs in Biomedicine*, vol. 156, pp. 25–45, 2018, doi: 10.1016/j.cmpb.2017.12.012.
- [13] A. T. Azar and S. M. El-Metwally, "Decision tree classifiers for automated medical diagnosis," *Neural Computing and Applications*, vol. 23, no. 7–8, pp. 2387–2403, 2013, doi: 10.1007/s00521-012-1196-7.
- [14] M. R. Senapati and P. K. Dash, "Local linear wavelet neural network based breast tumor classification using firefly algorithm," *Neural Computing and Applications*, vol. 22, no. 7–8, pp. 1591–1598, 2013, doi: 10.1007/s00521-012-0927-0.
- [15] A. T. Azar and S. A. El-Said, "Performance analysis of support vector machines classifiers in breast cancer mammography recognition," *Neural Computing and Applications*, vol. 24, no. 5, pp. 1163–1177, 2014, doi: 10.1007/s00521-012-1324-4.
- [16] E. Aličković and A. Subasi, "Breast cancer diagnosis using GA feature selection and rotation forest," *Neural Computing and Applications*, vol. 28, no. 4, pp. 753–763, 2017, doi: 10.1007/s00521-015-2103-9.
- [17] H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, "A support vector machine-based ensemble algorithm for breast cancer diagnosis," *European Journal of Operational Research*, vol. 267, no. 2, pp. 687–699, 2018, doi: 10.1016/j.ejor.2017.12.001.




- [18] V. Kumar, B. K. Mishra, M. Mazzara, D. N. H. Thanh, and A. Verma, "Prediction of malignant and benign breast cancer: a data mining approach in healthcare applications," *Lecture Notes on Data Engineering and Communications Technologies*, vol. 37, pp. 435–442, 2020, doi: 10.1007/978-981-15-0978-0\_43.
- [19] V. Lahoura *et al.*, "Cloud computing-based framework for breast cancer diagnosis using extreme learning machine," *Diagnostics*, vol. 11, no. 2, 2021, doi: 10.3390/diagnostics11020241.
- [20] M. Krzywinski and N. Altman, "Classification and regression trees," *Nature Methods*, vol. 14, no. 8, pp. 757–758, 2017, doi: 10.1038/nmeth.4370.
- [21] I. Vranckx, J. Raymaekers, B. De Ketelaere, P. J. Rousseeuw, and M. Hubert, "Real-time discriminant analysis in the presence of label and measurement noise," *Chemometrics and Intelligent Laboratory Systems*, vol. 208, 2021, doi: 10.1016/j.chemolab.2020.104197.
- [22] R. M. Mohana, C. K. K. Reddy, P. R. Anisha, and B. V. R. Murthy, "Random forest algorithms for the classification of tree-based ensemble," *Materials Today: Proceedings*, 2021, doi: 10.1016/j.matpr.2021.01.788.
- [23] C. Bowd *et al.*, "Gradient-boosting classifiers combining vessel density and tissue thickness measurements for classifying early to moderate glaucoma," *American Journal of Ophthalmology*, vol. 217, pp. 131–139, 2020, doi: 10.1016/j.ajo.2020.03.024.
- [24] J. Jorge and R. Paredes, "Passive-aggressive online learning with nonlinear embeddings," *Pattern Recognition*, vol. 79, pp. 162–171, 2018, doi: 10.1016/j.patcog.2018.01.019.
- [25] D. Martínez-Rego, D. Fernández-Francos, O. Fontenla-Romero, and A. Alonso-Betanzos, "Stream change detection via passive-aggressive classification and Bernoulli CUSUM," *Information Sciences*, vol. 305, pp. 130–145, 2015, doi: 10.1016/j.ins.2015.01.022.
- [26] Z. Khandezamin, M. Naderan, and M. J. Rashti, "Detection and classification of breast cancer using logistic regression feature selection and GMDH classifier," *Journal of Biomedical Informatics*, vol. 111, Nov. 2020, doi: 10.1016/j.jbi.2020.103591.
- [27] F. Morais-Rodrigues *et al.*, "Analysis of the microarray gene expression for breast cancer progression after the application modified logistic regression," *Gene*, vol. 726, 2020, doi: 10.1016/j.gene.2019.144168.
- [28] A. Bustamam, A. Bachtar, and D. Sarwinda, "Selecting features subsets based on support vector machine-recursive features elimination and one dimensional-naïve bayes classifier using support vector machines for classification of prostate and breast cancer," *Procedia Computer Science*, vol. 157, pp. 450–458, 2019, doi: 10.1016/j.procs.2019.08.238.
- [29] S. Cui *et al.*, "Using Naïve Bayes classifier to predict osteonecrosis of the femoral head with cannulated screw fixation," *Injury*, vol. 49, no. 10, pp. 1865–1870, 2018, doi: 10.1016/j.injury.2018.07.025.
- [30] E. N. Tamatjita and A. W. Mahastama, "Comparison of music genre classification using nearest centroid classifier and k-nearest Neighbours," *Proceedings of 2016 International Conference on Information Management and Technology, ICIMTech 2016*, pp. 118–123, 2017, doi: 10.1109/ICIMTech.2016.7930314.
- [31] F. Silva, M. Sanz, J. Seixas, E. Solano, and Y. Omar, "Perceptrons from memristors," *Neural Networks*, vol. 122, pp. 273–278, 2020, doi: 10.1016/j.neunet.2019.10.013.
- [32] B. A. Akinnuwesi, B. O. Macaulay, and B. S. Aribisala, "Breast cancer risk assessment and early diagnosis using principal component analysis and support vector machine techniques," *Informatics in Medicine Unlocked*, vol. 21, 2020, doi: 10.1016/j.imu.2020.100459.
- [33] W. H. Wolberg, O. Mangasarian, and D. W. Aha, "Breast Cancer Wisconsin (Original) Data Set," *UC Irvine Machine Learning Repository*, 1992. [https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+\(original\)](https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original)).

## BIOGRAPHIES OF AUTHORS



**Rania R. kadhim**    holds a Master of Science (M.Sc.) in physics from college of science, Mustansiriyah university, in 2016. She is currently a Ph.D. student in digital image processing at Mustansiriyah university in Iraq. Her research areas of interest include medical image processing, machine learning, computer vision, and image classification. She can be contacted at email: r90r61@gmail.com.



**Mohammed Y. Kamil**    holds a master of science (M.Sc.) in optics from Mustansiriyah university, in 2005. and Ph.D. in digital image processing from Mustansiriyah university, Iraq, in 2011, besides several professional certificates and skills. He is currently a professor with the physics department at Mustansiriyah university, Baghdad, Iraq. He is a member of the institute of electrical and electronics engineers (IEEE) and an editor in the Iraqi Journal of Physics. His research areas of interest include medical image processing, computer vision, and artificial intelligence. He can be contacted at email: m80y98@uomustansiriyah.edu.iq.